[0061] Figure 5 provides the results of an assay of GlycoPEGylation of EPO using the refolded SuperGlycoMix. Lanes are as follows: (1) MW markers, SeeBlue2 Invitrogen,(250, 148, 98, 64, 50, 36, 22, 16, 6 kD); (2) Positive control with EPO, + NSO expressed GalT1, BV GnT1, Aspergillus ST3GalIII and sugar nucleotides; (3) Negative control, Same as 2 without UDP-GlcNAc; (4) EPO, Purified and separately refolded MBP-GalT1(Δ129) C342T, Refolded MBP-GnT1(Δ103), and *Aspergillus niger* expressed ST3GalIII; (5) EPO, SuperGlycoMix (mixture of MBP-ST3GalIII, MBP-GalT1(Δ129) C342T, MBP-GnT1(Δ103)C123A and sugar nucleotides.

[0062] Figure 6 provides an alignment of a human GnT1 amino acid sequence (top line, NP_002397; SEQ ID NO:1) and a rabbit GnT1 amino acid sequence (bottom line, P27115; SEQ ID NO:2). The conserved unpaired cysteines are underlined and in bold text.

[0063] Figure 7 provides the amino acid sequence (SEQ ID NO:3) of a GnT1 Cys121Ser mutant and a nucleic acid sequence (SEQ ID NO:4) that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrrsldkllh.... (SEQ ID NO:5), where the bold residue is mutated from the wild-type cysteine.

[0064] Figure 8 provides the amino acid sequence (SEQ ID NO:6) of a GnT1 Cys121Asp mutant and a nucleic acid sequence (SEQ ID NO:7) that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrr**d**ldkllh... (SEQ ID NO:8), where the bold residue is mutated from the wild-type cysteine.

[0065] Figure 9 provides the amino acid sequence (SEQ ID NO:9) of a GnT1 Cys121Thr mutant and a nucleic acid sequence (SEQ ID NO:10) that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrr**t**ldkllh...(SEQ ID NO:11) , where the bold residue is mutated from the wild-type cysteine.

[0066] Figure 10 provides the amino acid sequence (SEQ ID NO:12) of a GnT1 Cys121Ala mutant and a nucleic acid sequence (SEQ ID NO:13) that encodes the mutant GnT1 protein.

The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrraldkllh... (SEQ ID NO:14), where the bold residue is mutated from the wild-type cysteine.

[0067] Figure 11 provides the amino acid sequence (SEQ ID NO:15) of a GnT1 Arg120Ala, Cys121His mutant and a nucleic acid sequence (SEQ ID NO:16) that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following double mutation: ...stvr**ah**ldkllh... (SEQ ID NO:17), where the bold residue is mutated from the wild-type cysteine.

[0068] Figure 12 provides the amino acid sequence of rat liver ST3GalIII (SEQ ID NO:18). The underlined and italicized sequence was deleted to make the Δ28 deletion.

[0069] Figures 13A and 13B provide full length nucleic acid (SEQ ID NO:20) and amino acid (SEQ ID NO:19) sequences of UDP-N-acetylgalactosaminyltransferase 2 (GalNAcT2). The accession number of the nucleic acid and protein is NM_004481.

[0070] Figures 14A and 14B provide nucleic acid (SEQ ID NO:22) and amino acid (SEQ ID NO:21) sequences of a Δ51GalNAcT2. The numbering is based on the full length amino acid and nucleic acid sequences shown in Figures 13A and B.

[0071] Figure 15 provides a demonstration of the protein concentration of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The pH values tested are expressed as solubilization pH-refolding pH. Protein concentrations were measured immediately after refolding (light gray bars), after dialysis (dark gray bars), and after concentration (white bars).

[0072] Figure 16 provides a demonstration of the enzymatic activity of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The pH values tested are expressed as solubilization pH-refolding pH. Activity was measured after dialysis (light gray bars) and after concentration (dark gray bars).

[0073] Figure 17 provides a demonstration of the specific activity of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The pH values tested are expressed as solubilization pH-refolding pH. Specific activity was measured after dialysis (white bars) and after concentration (dark gray bars).

[0074] Figures 18A and 18B provide results of remodeling of recombinant granulocyte colony stimulating factor (GCSF) using refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. Figure 18A shows the results using a control purified MBP-GalNAcT2(D51), or a negative control that lacked a substrate, or

bacterially expressed MBP-GalNAcT2(D51) that was solubilized at pH 6.5 and refolded at pH 6.5. Figure 18B shows the experimental results.

[0075] Figure 19 provides a profile of refolded MBP-GalNAcT2(D51) proteins after elution from a Q Sepharose XL (QXL) column (Amersham Biosciences, Piscataway, NJ). The top of the figure shows a chromatogram illustrating the elution of MBP-GalNAcT2(D51) from the QXL column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis. The bottom shows an image of two electrophoretic gels used to visualize the eluted fractions. The contents of each lane on the gel are described in the figure.

[0076] Figure 20 provides the GalNAcT2 activity of specific column fractions from the QXL column shown in Figure 19. The most active fractions were applied to a Hydroxyapatite Type I (80μm) (BioRad, Hercules, CA) column.

[0077] Figure 21 provides a profile of refolded MBP-GalNAcT2(D51) proteins after elution from the HA type I column. The top of the figure shows a chromatogram illustrating the elution of MBP-GalNAcT2(D51) from the HA type I column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis. The bottom shows an image of an electrophoretic gel used to visualize the eluted fractions. The contents of each lane on the gel are described in the figure.

[0078] Figure 22 provides the GalNAcT2 activity of HA type I eluted fractions.

[0079] Figure 23 provides a comparison of purification and activity of ST3Gal3 proteins fused to either an MBP tag or to an MBP-SBD tag.

[0080] Figure 24 provides the amino acid sequences of the MBP-ST3Gal1 fusion protein (SEQ ID NO:23) (A) and the MBP-SBD-ST3Gal1 fusion protein (SEQ ID NO:24) (B).

[0081] Figure 25 provides the sialyltransferase activity of the MBP-ST3Gal3 fusion protein) and the MBP-SBD-ST3Gal3 fusion protein. positive and negative controls are also shown.

[0082] Figure 26 provides the amino acid sequence of mouse and human ST6GalNAcI proteins fused to MBP. Part A shows the sequence of a mouse truncation fusion: MBP-mST6GalNAcI S127 (SEQ ID NO:25). Part B shows the sequence of a human truncation fusion: MBP-hST6GalNAcI K36 (SEQ ID NO:26).

[0083] Figure 27 provides SDS-PAGE gels of O-linked glycosyltransferase enzyme (A) concentrations after co-refolding and the (B) results of an enzyme assay after co-refolding. MBP-GalNAcT2 and MBP-ST3GalI were co-refolded together. Enzyme activity was tested after addition of Core I Gal T1 enzyme. The substrates were IFα-2b and 20K-Peg-CMP-NAN.

[0084] Figure 28 provides an SDS-PAGE gel showing expression of the native SiaA protein in *E. coli* before and after induction with IPTG.

[0085] Figure 29 provides an SDS-PAGE gel showing expression of an MBP-SiaA fusion protein in *E. coli* before and after induction with IPTG.

[0086] Figure 30 provides the amino acid sequence of the full length bovine GalT1 protein (SEQ ID NO:27).

[0087] Figure 31 depicts GalT1 mutants schematically, as well as a control protein GalT1(40) (S96A+C342T).

[0088] Figure 32 provides the results of enzymatic assays of the refolded and purified MBP-GalT1 (D70) protein. The assay measured conversion of LNT2 (Lacto-N-Triose-2) into LNnT (Lacto-N-Neotetraose) using UDP-Gal (Uridine 5'-Diphosphogalactose) as a donor substrate.

[0089] Figure 33 provides an RNAse B remodeling assay of MBP-GalT1 (D70) and a control protein GalT1(40) (S96A+C342T), also referred to as Qasba's GalT1.

[0090] Figure 34 provides kinetics of glycosylation of RNAse B using the refolded and purified MBP-GalT1 (D70) protein or NSO GalT1, a soluble form of the bovine GalT1 protein that was expressed in a mammalian cell system.

[0091] Figure 35 provides a schematic of the MBP-GnT1 fusion proteins, and depicts the truncations, *e.g.*, Δ103 or Δ35, and the Cys121Ser mutation (top). The bottom of the figure provides the full length human GnT1 protein (SEQ ID NO:1).

[0092] Figure 36 provides an SDS-PAGE gel showing in the right panel the refolded MBP-GnT1 fusion proteins: MBP-GnT1(D35) C121A, MBP-GnT1(D103) R120A + C121H, and MBP-GnT1(D103) C121A. The left panel shows GnT1 activities of two different batches (A1 and A2) of refolded MBP-GnT1(D35) C121A at different time points.

[0093] Figure 37 provides a full length sequence of porcine ST3Gal1 (SEQ ID NO:28).

[0094] Figure 38 provides full length amino acid sequences for A) human ST6GalNAcTI (SEQ ID NO:29) and for B) chicken ST6GalNAcTI (SEQ ID NO:30), and C) a sequence of the mouse ST6GalNAcTI protein beginning at residue 32 of the native mouse protein (SEQ ID NO:31).

[0095] Figure 39 provides a schematic of a number of preferred human ST6GalNAcI truncation mutants.

[0096] Figure 40 shows a schematic of MBP fusion proteins including the human ST6GalNAcI truncation mutants.

[0097] Figure 41 provides the full length sequence of human Core 1 GalT1 protein (SEQ ID NO:32).

[0098] Figure 42 provides the sequences of two Drosophila Core 1 GalT1 proteins (SEQ ID NOS:33 and 34).

[0099] Figure 43 provides the sequences of exemplary bacterial MBP proteins that can be fused to glycosyltransferases to enhance refolding. A. *Yersinia* MBP (SEQ ID NO:35); B. *E. coli* MBP (SEQ ID NO:36); C. *Pyrococcus furiosus* MBP (SEQ ID NO:37); D. *Thermococcus litoralis* MBP (SEQ ID NO:38); E. *Thermatoga maritime* MBP (SEQ ID NO:39); and F. *Vibrio cholerae* MBP (SEQ ID NO:40).

[0100] Figure 44 provides an alignment of human GalNAcT1 (SEQ ID NO:41) and GalNAcT2 proteins (SEQ ID NO:19). Because the alignment programs account for sequence insertions or deletions, the numbering of cysteine residues is not the same as mentioned text and published sequences. In the case of hGalNAc-T2 cysteine 227 (published) corresponds to position 235 in the alignment and cysteine 229 (published) is 237 in the alignment. The hGalNAc-T1 cysteines are 212 (published), which corresponds to cysteine 235 (alignment) and 214 (published), which corresponds to cysteine 237 (alignment). The relevant cysteine residues are indicated by larger font size. Consensus peptides = SEQ ID NOS:42-65.

[0101] Figure 45 shows the position of paired and unpaired cysteine residues in the human ST6GalNAcI protein. Single and double cysteine substitution are also shown, *e.g.*, C280S, C362S, C362T, (C280S + C362S), and (C280S + C362T).

DEFINITIONS

[0102] The recombinant glycosyltransferase proteins of the invention are useful for transferring a saccharide from a donor substrate to an acceptor substrate. The addition generally takes place at the non-reducing end of an oligosaccharide or carbohydrate moiety on a biomolecule. Biomolecules as defined here include but are not limited to biologically

[0133] A "fusion protein" refers to a protein comprising amino acid sequences that are in addition to, in place of, less than, and/or different from the amino acid sequences encoding the original or native full-length protein or subsequences thereof.

[0134] Components of fusion proteins include "accessory enzymes" and/or "purification tags." An "accessory enzyme" as referred to herein, is an enzyme that is involved in catalyzing a reaction that, for example, forms a substrate for a glycosyltransferase. An accessory enzyme can, for example, catalyze the formation of a nucleotide sugar that is used as a donor moiety by a glycosyltransferase. An accessory enzyme can also be one that is used in the generation of a nucleotide triphosphate required for formation of a nucleotide sugar, or in the generation of the sugar which is incorporated into the nucleotide sugar. The recombinant fusion protein of the invention can be constructed and expressed as a fusion protein with a molecular "purification tag" at one end, which facilitates purification of the protein. Such tags can also be used for immobilization of a protein of interest during the glycosylation reaction. Suitable tags include "epitope tags," which are a protein sequence that is specifically recognized by an antibody. Epitope tags are generally incorporated into fusion proteins to enable the use of a readily available antibody to unambiguously detect or isolate the fusion protein. A "FLAG tag" is a commonly used epitope tag, specifically recognized by a monoclonal anti-FLAG antibody, consisting of the sequence AspTyrLysAspAspAspAspLys (SEQ ID NO:66) or a substantially identical variant thereof. Other suitable tags are known to those of skill in the art, and include, for example, an affinity tag such as a hexahistidine peptide (SEQ ID NO:67), which will bind to metal ions such as nickel or cobalt ions. Proteins comprising purification tags can be purified using a binding partner that binds the purification tag, *e.g.*, antibodies to the purification tag, nickel or cobalt ions or resins, and amylose, maltose, or a cyclodextrin. Purification tags also include starch binding domains, *E. coli* thioredoxin domains (vectors and antibodies commercially available from *e.g.*, Santa Cruz Biotechnology, Inc. and Alpha Diagnostic International, Inc.), and the carboxy-terminal half of the SUMO protein (vectors and antibodies commercially available from *e.g.*, Life Sensors Inc.). Maltose binding domains are preferably used for their ability to enhance refolding of insoluble eukaryotic glycosyltransferases, but can also be used to assist in purification of a fusion protein. Purification of maltose binding domain proteins is known to those of skill in the art. Starch binding domains are described in WO 99/15636, herein incorporated by reference. Affinity purification of a fusion protein comprising a starch binding domain using

about amino acid residues 32-90. Thus, a truncated human Core1 GalT1 protein can have deletions at the amino terminus of about *e.g.*, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90 residues.

[0210] Deletion mutations can also be made in an ST3Gal1 protein. For example, the human ST3Gal1 protein includes a stem region from about amino acid residues 18-58. Thus, a truncated human ST3Gal1 protein can have deletions at the amino terminus of about *e.g.*, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, or 58 residues. As another example, the porcine ST3Gal1 protein includes a stem region from about amino acid residues 28-61. Thus, a truncated porcine ST3Gal1 protein can have deletions at the amino terminus of about *e.g.*, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, or 61 residues.

[0211] Deletion mutations can also be made in a GalNAcT2 protein. For example, the rat GalNAcT2 protein includes a stem region from about amino acid residues 40-95. Thus, a truncated rat GalNAcT2 protein can have deletions at the amino terminus of about *e.g.*, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, or 95 residues.

[0212] Deletion mutations can also be made in an ST6GalNAcI protein. For example, the mouse ST6GalNAcI protein includes a stem region from about amino acid residues 30-207. Thus, a truncated mouse ST6GalNAcI protein can have deletions at the amino terminus of about *e.g.*, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197 198, 199, 200, 201, 202, 203, 204, 205, 206, or 207 residues. As another example, the human ST6GalNAcI protein includes a stem region from about amino

acid residues 35-278. Thus, a truncated human ST6GalNAcI protein can have deletions at the amino terminus of about *e.g.*, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277,or 278 residues. As still another example, chicken ST6GalNAcI protein includes a stem region from about amino acid residues 37-253. Thus, a truncated chicken ST6GalNAcI protein can have deletions at the amino terminus of about *e.g.*, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, or 253 residues.

D. One pot refolding of glycosyltransferases

[0213] These embodiments of the invention are based on the surprising observation that multiple eukaryotic glycosyltransferases expressed in bacterial inclusion bodies can be refolded in a single vessel, *i.e.*, a one pot method. Using this method at least two glycosyltransferases can be refolded together resulting in savings of time and materials.

residues of the yeast ubiquitin gene containing a peptidase cleavage site. Cleavage at the junction of the two moieties results in production of a protein having an intact authentic N-terminal reside.

[0259] The expression vectors of the invention can be transferred into the chosen host cell by well-known methods such as calcium chloride transformation for *E. coli* and calcium phosphate treatment or electroporation for mammalian cells. Cells transformed by the plasmids can be selected by resistance to antibiotics conferred by genes contained on the plasmids, such as the *amp, gpt, neo* and *hyg* genes.

VI. Proteins and protein purification

[0260] The recombinant eukaryotic glycosyltransferase proteins can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (*see*, generally, R. Scopes, *Protein Purification*, Springer-Verlag, N.Y. (1982), Deutscher, *Methods in Enzymology Vol. 182: Guide to Protein Purification.*, Academic Press, Inc. N.Y. (1990)). In preferred embodiments, purification of the recombinant eukaryotic glycosyltransferase proteins occurs after refolding of the protein. Substantially pure compositions of at least about 70 to 90%, homogeneity are preferred; more preferably at least 91%, 92%, 93%, 94%, 95%, 96%, or 97%; and 98 to 99% or more homogeneity are most preferred. The purified proteins may also be used, *e.g.*, as immunogens for antibody production.

[0261] To facilitate purification of the recombinant eukaryotic glycosyltransferase proteins of the invention, the nucleic acids that encode the recombinant eukaryotic glycosyltransferase proteins can also include a coding sequence for an epitope or "tag" for which an affinity binding reagent is available, *i.e.* a purification tag. Examples of suitable epitopes include the myc and V-5 reporter genes; expression vectors useful for recombinant production of fusion proteins having these epitopes are commercially available (*e.g.*, Invitrogen (Carlsbad CA) vectors pcDNA3.1/Myc-His and pcDNA3.1/V5-His are suitable for expression in mammalian cells). Additional expression vectors suitable for attaching a tag to the fusion proteins of the invention, and corresponding detection systems are known to those of skill in the art, and several are commercially available (e.g., "FLAG" (Kodak, Rochester NY). Another example of a suitable tag is a polyhistidine sequence, which is capable of binding to metal chelate affinity ligands. Typically, six adjacent histidines (SEQ ID NO:67) are used,

although one can use more or less than six. Suitable metal chelate affinity ligands that can serve as the binding

enzyme, GST-ST3-GalIII, was active and transferred sialic acid to an LNnT sugar substrate and to asialylated glycoproteins, for example, transferrin and Factor IX.

Cloning ST3GalIII into pGEX-XT-KT vector

[0305] Rat liver ST3-GalIII gene was cloned into *Bam*H1 and *Eco*R1 sites of the pGEX-KT-Ext vector after PCR Amplification using the following primers:

Sense Sial 5'Tm 5'-TTTGGATCCAAGCTACACTTACTCCAATGG (SEQ ID NO:68)
Antisense: Sial 3' Whole 5'-TTTGAATTCTCAGATACCACTGCTTAAGTC (SEQ ID NO:69)

Expression of GST-ST3GalIII in *E. coli* BL21cells

[0306] pGEX-ST3GalIII, an expression vector comprising the ST3GalIII GST fusion, was transformed into chemically competent *E. coli* BL21 cells. Single colonies were picked, inoculated into five ml LB media with 100 μg/ml carbenicillin, and grown overnight at 37°C with shaking. The next day, one ml of overnight culture was transferred into one liter of LB media with 100 μg/ml carbenicillin. Bacteria were grown until to an OD_{620} of 0.7, then 150 μM IPTG (final) was added to the medium. Bacteria were grown at 37°C for one to two hours more, then shifted to room temperature and grown overnight with shaking. Cells were harvested by centrifugation; bacterial pellets were resuspended in PBS buffer and lysed using a French Press. Soluble and insoluble fractions were separated by centrifugation for thirty minutes at 10,000 RPM in a Sorvall, SS 34 rotor at 4°C.

Purification of the inclusion bodies

[0307] Fifty ml of Novagen's Wash buffer (20 mM Tris.HCl, pH 7.5, 10 mM EDTA, 1 % Triton X-100) was added to the insoluble fraction, *i.e.*, the inclusion bodies (IB's). The insoluble fraction was vortexed to resuspend the pellet. The suspended IB's were centrifuged and washed at least twice by resuspending in Wash Buffer as above. Clean precipitates (IB's) were recovered and were stored at -20 °C until use.

Refolding inclusion bodies

[0308] The IB's were weighed (144 mg) and dissolved in Genotech IBS buffer (1.44 ml). The resuspended IB's were incubated at 4 °C for one hour in an Eppendorf centrifuge tube. Insoluble material was removed by centrifugation at maximum speed in an Eppendorf centrifuge. Solubilized IB's were diluted to 4 ml final volume. Refolding of GST-ST3GalIII was tested in refolding buffer solutions containing cyclodextrin, polyethylene glycol (PEG),

ND SB-201, or a GSH/GSSG redox system. One ml of solubilized IB's were diluted rapidly by pipetting into the refolding solution, vigorously mixed for 30-40 seconds, and then gently

Table 3. GST-ST3GalIII activities after two separate folding experiments using GSH/GSSG system.

GSH/GSSG	Conc	Activity
Refolding Trial 1	12x	182 U/L*
Refolding Trial 2	40x	531 U/L*

*Activities reported here are Units per L refolded enzyme

Sialylation of glycoproteins using refolded GST-ST3 Gal III

[0311] Twenty μL of asialylated Transferrin (2μg/μL) or asialylated Factor IX (2μg/μL), was added to fifty μL of a buffer containing 50mM Tris, pH 8.0; and 150 mM NaCl, with 10 μL of 100 mM $MnCl_2$; 10 μL of 200mM CMP-NAN; and 0.05% sodium azide. The reaction mixture was incubated with 30 μL refolded GST-ST3GalIII at 30°C overnight or longer with shaking at 250 rpm. After the reactions were stopped, the sialylated proteins were separated on pH 7-3 IEF (Isoelectric focusing gel, Invitrogen) and stained with Comassie Blue according to manufacturer's guideline. Both Transferrin and Factor IX were sialylated by GST-ST3GalIII. (Data not shown.)

Refolding a rat liver ST3GalIII fused to an MBP tag.

[0312] Rat liver ST3GalIII was cloned into pMAL-c2x vector and expressed as a maltose binding protein (MBP) fusion, MBP-ST3GalIII, in inclusion bodies of *E.coli* TB1 cells. The refolded MBP-ST3GalIII was active and transferred sialic acid to LNnT, a sugar substrate, and to asialylated glycoproteins, for example asialo-transferrin.

Cloning ST3GalIII into pMAL-c2x vector

[0313] The rat liver ST3-GalIII nucleic acid was cloned into *Bam*H1 and *Xba*I sites of the pMAL-c2x vector after PCR Amplification using the following primers:

Sense ST3BAMH1 5'-TAATGGATTCAAGCTACACTTACTCCAATGG (SEQ ID NO:70)
Antisense: ST3XBA1 5'-GCGCTCTAGATCAGATACCACTGCTTAAGT (SEQ ID NO:71)

[0314] Nucleotides encoding amino acids 28-374, *e.g.*, the stem region and catalytic domain of ST3GalIII, were fused to the MBP amino acid tag.

[0315] Three other truncations of ST3GalIII were constructed and fused to MBP. The three ST3Gal III (Δ73, Δ85, Δ86) inserts were isolated by PCR using the following 5'

primers (ST3 BamH1 Δ73) TGTATCGGATCCCTGGCCACCAAGTACGCTAACTT (SEQ ID NO:72); (ST3 BamH1 Δ85) TGTATCGGATCCTGCAAACCCGGCTACGCTTCAGCCAT (SEQ ID NO:73); and (ST3

BamH1 Δ86) TGTATCGGATCCAAACCCGGCTACGCTTCAGCCAT (SEQ ID NO:74) respectively, in pairs with the common 3' primer (ST3-Xho1)GGTCTCCTCGAGTCAGATACCACTGCTTAA (SEQ ID NO:75). Each PCR product was digested with BamHI and XhoI, subcloned into BamHI-XhoI digested pCWin2-MBP Kanr vector, transformed into TB1 cells, and screened for the correct construct.

[0316] PCR reactions were carried out under the following conditions. One cycle at 95°C for 1 minute. One μl vent polymerase was added. Ten of the following cycles were performed: 94°C for 1 minute; 65°C for 1 minute; and 72°C for 1 minute. After a final ten minutes at 72°C, the reaction was cooled to 4°C.

[0317] All of the ST3GalIII truncations had activity after refolding. The experiments described below were performed using the MBP Δ73ST3GalIII truncation.

Expression of MBP-ST3GalIII in *E. coli* TB1 cells

[0318] The pMAL-ST3GalIII plasmid was transformed into chemically competent *E. coli* TB1 cells. Three isolated colonies containing TB1/pMAL-ST3GalIII construct were picked from the LB agar plates. The colonies were grown in five ml of LB media supplemented with 60 μg/ml carbenicillin at 37°C with shaking until the liquid cultures reached an OD_{620} of 0.7. Two one ml aliquots were withdrawn from each culture and used to inoculate fresh media with or without 500 μM IPTG (final). The cultures were grown at 37°C for two hours. Bacterial cells were harvested by centrifugation. Total cell lysates were prepared heating the cell pellets in the presence of SDS and DTT. IPTG induced expression of MBP-ST3GalIII. (Data not shown.)

Expression of MBP-ST3GalIII and Purification of the inclusion bodies:

[0319] A one ml aliquot of TB1/pMAL-ST3GalIII overnight culture was inoculated into 0.5 liter of LB media with 50 μg/ml carbenicillin and grown to an OD_{620} of 0.7. Expression of MBP-ST3GalIII was induced by addition of 0.5 mM IPTG, followed by overnight incubation at room temperature. The next day bacterial cells were harvested by centrifugation. Cell pellets were resuspended in a buffer containing 75 mM TrisHCl, pH 7.4; 100 mM NaCl; and 1 % glycerol. Bacterial cells were lyzed using a French Press. Soluble and insoluble fractions were separated by centrifugation for thirty minutes, 4°C, 10,000 rpm, Sorvall, SS 34 rotor). Soluble and insoluble fractions were separated by centrifugation for thirty minutes at 10,000RPM in a Sorvall, SS 34 rotor at 4°C.

buffer was supplemented with 0.3 mM Lauryl maltoside (LM); 0.1 mM oxidized glutathione (GSSG); 1 mM reduced glutathione (GSH) immediately before the addition of solubilized IB's. Two ml of solubilized IB's were added into 43 ml of refolding buffer in 50 ml sterile culture tube. The tube was placed on a rocker-shaker and gently shaken for 24 hours at 4°C. The refolded protein was dialyzed in dialysis tubing (MWCO: 7 kD) against Dialysis Buffer (100 mM Tris HCl, pH 7.5; 100 mM NaCl; and 5 % glycerol) twice (in 10-20 volume excess buffer).

[0330] The large scale dialyzed, refolded MBP-Gal III was analyzed for ST3GalIII activity, and exhibited about 53.6 U/g IB.

Example 2: Site Directed Mutagenesis of Human GnTI to Enhance Refolding.

[0331] A truncated human N-acetylglucosaminyltransferase I (103 amino terminal amino acids deleted) was expressed in *E.coli* as a maltose binding fusion protein (GnTI/MBP). The fusion protein was insoluble and was expressed in inclusion bodies. After solubilization and refolding, the GnTI/MBP fusion protein had low activity. The crystal structure of a truncated form of rabbit GnTI (105 amino terminal amino acids deleted) shows an unpaired cysteine residue (CYS123) near the active site. (See, *e.g.*, Unligil *et al.*, *EMBO J.* 19:5269-5280 (2000)). The corresponding unpaired cysteine in the human GnTI was identified as CYS121 and was replaced with a series of amino acids that are similar in size and chemical characteristics. The amino acids used include serine (Ser), threonine (Thr), alanine (Ala) and aspartic acid (Asp). In addition, a double mutant, ARG120ALA, CYS121HIS, was also made. The mutant GnTI/MBP fusion proteins were expressed in *E. coli*, refolded and assayed for GnTI activity towards glycoproteins.

[0332] Mutagenesis was done using a Quick Change Site-Directed Mutagenesis Kit from Stratagene. Additional restriction sites were introduced with some of the GnT1 mutations. For example an *Apa*I site (underlined, GG**GCCCAC**) was introduced into the GnT1 ARG120ALA, CYS121HIS mutant, *i.e.*, CGC CTG → **GCC CAC** (changes in bold). The following mutagenic oligonucleotides were used to make the double mutant: GnT1 R120A, C121H+, 5'CCGCAGCACTGTTCGG**GCCCAC**CTGGACAAGCTGCTG 3' (SEQ ID NO:76); and GnT1 R120A, C121H-
5'CAGCAGCTTGTCCAG**GTGGGCCC**GAACAGTGCTGCGG 3' (SEQ ID NO:77) (changes shown in bold). An *Asc*I site (underlined, GGCGC**GCC**) was introduced into the

GnT1 CYS121ALA mutant, *i.e.*, CTG → **GCC** (changes in bold). The following mutagenic oligonucleotides were used to make the GnT1 CYS121ALA mutant: GnT1C123A+

5'AGCACTGTTCGGCGCGCCCTGGACAAGCTGCTG 3' (SEQ ID NO:78); and GnT1C123A-5'CAGCAGCTTGTCCAGGGCGCGCCGAACAGTGCT 3' (SEQ ID NO:79).

[0333] The activity of the mutant proteins expressed in *E. coli* was compared to the activity of wild type GnT1 expressed in baculovirus. A CYS121SER GNTI mutant was active in a TLC based assay. In contrast, a CYS121THR mutant had no detectable activity and a CYS121ASP mutant had low activity. A CYS121ALA mutant was very active, and a double mutant, ARG120ALA, CYS121HIS, based on the amino acid sequence of the *C. elegans* GnT1 protein (Gly14), also exhibited activity, including transfer of GlcNAc to glycoproteins. Amino acid and encoding nucleic acid sequences of the GnT1 mutants are provided in Figures 7-11.

[0334] A second GnT1 truncation was made and fused to MBP: MBP-GnT1(D35). Figure 35 provides a schematic of the MBP-GnT1 fusion proteins, and depicts the truncations, *e.g.*, Δ103 or Δ35, and the Cys121Ser mutation (top). The bottom of the figure provides the full length human GnT1 protein. Mutations of Cys121 were also made in the MBP-GnT1(D35) protein.

[0335] Both fusion proteins were expressed in E. coli and both had activity for remodeling of the RNAse B glycoprotein. Figure 36 provides an SDS-PAGE gel showing in the right panel the refolded MBP-GnT1 fusion proteins: MBP-GnT1(D35) C121A, MBP-GnT1(D103) R120A + C121H, and MBP-GnT1(D103) C121A. The left panel shows the activities for remodeling the RNAse B glycoprotein of two different batches (A1 and A2) of refolded MBP-GnT1(D35) C121A at different time points. The MBP-GnT1 (D103) C121A also remodeled the RNAse B glycoprotein. Data not shown.

Example 3: MPB fusions to GalT1.

[0336] The following fusions between truncated bovine GalT1 and MBP were constructed: MBP-GalT1 (D129) wt, (D70) wt or (D129 C342T). (For the full length bovine sequence, see, *e.g.*, D'Agostaro *et al.*, *Eur. J. Biochem.* 183:211-217 (1989) and accession number CAA32695.) Each construct had activity after refolding. The amino acid sequence of the full length bovine GalT1 protein is provided in Figure 30. The mutants are depicted shematically in Figure 31 with a control protein GalT1(40) (S96A+C342T). *See, e.g.*, Ramakrishnan *et al.*, *J. Biol. Chem.* 276:37666-37671 (2001).

[0337] MBP-GalT1 (D70) was expressed in *E. coli* strain JM109. After overnight induction with IPTG, inclusion bodies were isolated from the insoluble pellet after cells were

Example 5: Refolding eukaryotic GalNAcT2.

[0365] A truncated human GalNAcT2 enzyme was expressed in *E. coli* and used to determine optimal conditions for solubilization and refolding using the methods described above. The full length human GalNAcT2 nucleic acid and amino acid sequences are provided in Figures 13A and B. The sequences of the mutant protein, GalNAcT2(D51), are shown in Figures 14A and B. The mutant was expressed in *E. coli* as an MBP fusion protein, MBP-GalNAcT2(D51). Other GalNAcT2 mutants were made, expressed in E. coli and were able to be refolded: MBP-GalNAcT2(D40), MBP-GalNAcT2(D73), and MBP-GalNAcT2(D94). Data not shown. Details of the construction of the additional deletion mutants is found in USSN 60/576,530, filed June 3, 2004 and USSN 60/598,584, August 3, 2004, both of which are herein incorporated by reference for all purposes.

[0366] Cultures of bacteria expressing MBP-GalNAcT2(D51) were grown and harvested as described above. Inclusion bodies were purified from bacteria as described above. Solubilization of the inclusion bodies was performed at pH 6.5 or at pH 8.0. After solubilization, MBP-GalNAcT2(D51) protein was refolded at either pH 6.5 or pH 8.0 using buffers A and B, *i.e.*, **Buffer A**: 55 mM MES pH 6.5, 550 mM Arginine, 0.055 % PEG3350, 264 mM NaCl, 11 mM KCl, supplemented with 1 mM GSH, 0.1 mM GSSG; and **Buffer B**: 55 mM TrisHCl pH 8, 550 mM Arginine, 0.055 % PEG3350, 264 mM NaCl, 11 mM KCl, supplemented with 1 mM GSH, 0.1 mM GSSG. After refolding, MBP-GalNAcT2(D51) protein was dialyzed and then concentrated. Figure 15 provides a demonstration of the protein concentration of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0.

[0367] A radiolabeled [^{3}H]-UDP-GalNAc assay was performed to determine the activity of the *E.coli*-expressed refolded MBP-GalNAcT2(D51) by monitoring the addition of radiolabeled GalNAc to a peptide acceptor. The acceptor was a MuC-2 – like peptide having the sequence MVTPTPTPTC (SEQ ID NO:80). The peptide was dissolved in 1M Tris-HCl pH=8.0. See, *e.g.*, USSN 60/576,530 filed June 3, 2004; and US provisional patent application Attorney Docket Number 040853-01-5149-P1, filed August 3, 2004; both of which are herein incorporated by reference for all purposes. Figure 16 provides a demonstration of the enzymatic activity of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. Figure 17 provides a demonstration of the specific activity of refolded MBP-GalNAcT2(D51) after solubilization

at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The highest activity levels were observed with MBP-GalNAcT2(D51) that

Human ST6GalNAcTI
MRSCLWRCRHLSQGVQWSLLLAVLVFFLFALPSFIKEPQTKPSRHQRTENIKERSLQS
LAKPKSQAPTRARRTTIYAEPVPENNALNTQTQPKAHTTGDRGKEANQAPPEEQDK
VPHTAQRAAWKSPEKEKTMVNTLSPRGQDAGMASGRTEAQSWKSQDTKTTQGNG
GQTRKLTASRTVSEKHQGKAATTAKTLIPKSQHRMLAPTGAVSTRTRQKGVTTAVIP
PKEKKPQATPPPAPFQSPTTQRNQRLKAANFKSEPRWDFEEKYSFEIGGLQTTCPDSV
KIKASKSLWLQKLFLPNLTLFLDSRHFNQSEWDRLEHFAPPFGFMELNYSLVQKVVT
RFPPVPQQQLLLASLPAGSLRCITCAVVGNGGILNNSHMGQEIDSHDYVFRLSGALIK
GYEQDVGTRTSFYGFTAFSLTQSLLILGNRGFKNVPLGKDVRYLHFLEGTRDYEWLE
ALLMNQTVMSKNLFWFRHRPQEAFREALHMDRYLLLHPDFLRYMKNRFLRSKTLD
GAHWRIYRPTTGALLLLTALQLCDQVSAYGFITEGHERFSDHYYDTSWKRLIFYINH
DFKLEREVWKRLHDEGIIRLYQRPGPGTAKAKN

FIG. 38A

Chicken ST6GalNAcTI
MGFLIRRLPKDSRIFRWLLILTVFSFIITSFSALFGMEKSIFRQLKIYQSIAHMLQVDTQ
DQQGSNYSANGRISKVGLERDIAWLELNTAVSTPSGEGKEEQKKTVKPVAKVEEAK
EKVTVKPFPEVMGITNTTASTASVVERTKEKTTARPVPGVGEADGKRTTIALPSMKE
DKEKATVKPSFGMKVAHANSTSKDKPKAEEPPASVKAIRPVTQAATVTEKKKLRAA
DFKTEPQWDFDDEYILDSSSPVSTCSESVRAKAAKSDWLRDLFLPNITLFIDKSYFNV
SEWDRLEHFAPPYGFMELNYSLVEEVMSRLPPNPHQQLLLANSSSNVSTCISCAVVG
NGGILNNSGMGQEIDSHDYVFRVSGAVIKGYEKDVGTKTSFYGFTAYSLVSSLQNLG
HKGFKKIPQGKHIRYIHFLEAVRDYEWLKALLLDKDIRKGFLNYYGRRPRERFDEDF
TMNKYLVAHPDFLRYLKNRFLKSKNLQKPYWRLYRPTTGALLLLTALHLCDRVSAY
GYITEGHQKYSDHYYDKEWKRLVFYVNHDFNLEKQVWKRLHDENIMKLYQRS

FIG. 38B

Mouse ST6GalNAcTI protein beginning at residue 32 of the native mouse protein
DPRAKDSRCQFIWKNDASAQENQQKAEPQVPIMTLSPRVHNKESTSVSSKDLKKQER
EAVQGEQAEGKEKRKLETIRPAPENPQSKAEPAAKTPVSEHLDKLPRTPGALSTRKTP
MATGAVPAKKKVVQATKSPASSPHPTTRRRQRLKASEFKSEPRWDFEEEYSLDMSSL
QTNCSASVKIKASKSPWLQNIFLPNITLFLDSGRFTQSEWNRLEHFAPPFGFMELNQSL
VQKVVTRFPPVRQQQLLLASLPTGYSKCITCAVVGNGGILNDSRVGREIDSHDYVFR
LSGAVIKGYEQDVGTRTSFYGFTAFSLTQSILILGRRGFQHVPLGKDVRYLHFLEGTR
NYEWLEAMFLNQTLAKTHLSWFRHRPQEAFRNALDLDRYLLLHPDFLRYMKNRFL
RSKTLDTAHWRIYRPTTGALLLLTALHLCDKVSAYGFITEGHQRFSDHYYDTSWKRL
IFYINHDFRLERMVWKRLHDEGIIWLYQRPQSDKAKN

FIG. 38C

```
                    1                                                  50
HSGALNAT1.pep     (1)MRKFAYCKVVLATSLIWVLLDMFLLLYFS----------------ECNKC
HSGALNAT2.pep     (1)MRRRS--RMLLCFAFLWVLGIAYYMYSGGGSALAGGAGGGAGRKEDWNEI
    Consensus     (1)MRK A  KMLL  A IWVL   F L                    D N
                    51                                                 100
HSGALNAT1.pep    (35)DEKKERGLPAGDVLEPVQKPHEGP-G-----EMGKPVVIPKEDQEKMKEM
HSGALNAT2.pep    (49)DPIKKKDLHHSNGEEKAQSMETLPPGKVRWPDFNQEAYVGGIMVRSGQDP
    Consensus    (51)D  K K L      E  Q     P G     D      I         D
                   101                                                 150
HSGALNAT1.pep    (79)FKINQFNLMASEMIALNRSLPDVRLEGCKTKVYPDNLPTTSVVIVFHNEA
HSGALNAT2.pep    (99)YARNKFNQVESDKLRMDRAIPDTRHDQCQRKQWRVDLPATSVVITFHNEA
    Consensus   (101)F  N FN M SD I L RAIPD R D C  K W   LP TSVVI FHNEA
                   151                                                 200
HSGALNAT1.pep   (129)WSTLLRTVHSVINRSPRHMIEEIVLVDDASERDFLKRPLESYVKKLKVPV
HSGALNAT2.pep   (149)RSALLRTVVSVLKKSPPHLIKEIILVDDYSN-----DPEDGALLGKIEKV
    Consensus   (151) S LLRTV SVI KSP HLI EIILVDD S       P D  L      V
                   201                                                 250
HSGALNAT1.pep   (179)HVIRMEQRSGLIRARLKGAAVSKGQVITFLDAHCECTVGWLEPLLARIKH
HSGALNAT2.pep   (194)RVLRNDRREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAE
    Consensus   (201)  VIR D R GLIRARLKGA  A A VITFLDAHCEC   WLEPLL RI
                   251                                                 300
HSGALNAT1.pep   (229)DRRTVVCPIIDVISDDTFEYMAGSDMTYGGFNWKLNFRWYPVPQREMDRR
HSGALNAT2.pep   (244)DRTRVVSPIIDVINMDNFQYVGASADLKGGFDWNLVFKWDYMTPEQRRSR
    Consensus   (251)DR  VV PIIDVI  D F YMAAS    GGF W L FKW  M       R
                   301                                                 350
HSGALNAT1.pep   (279)KGDRTLPVRTPIMAGGLFSIDRDYFQEIGTYDAGMDIWGGENLEISFRIW
HSGALNAT2.pep   (294)QGNPVAPIKTPMIAGGLFVMDKFYFEELGKYDMMMDVWGGENLEISFRVW
    Consensus   (301)  G    PIKTP IAGGLF IDK YF EIG YD  MDIWGGENLEISFRIW
                   351                                                 400
HSGALNAT1.pep   (329)QCGGTLEIVTCSHVGHVFRKATPYTFPGGIGQIINKNNRRLAEVWMDEFK
HSGALNAT2.pep   (344)QCGGSLEIIPCSRVGHVFRKQHPYTFPGGSGIVFARNTRRAAEVWMDEYK
    Consensus   (351)QCGGSLEII CS VGHVFRK  PYTFPGGSG I  KN RR AEVWMDEFK
                   401                                                 450
HSGALNAT1.pep   (379)NFFYIISPGVTKVDYGDISSRVGLRHKLQCKPFSWYLENIYPDSQIPRHY
HSGALNAT2.pep   (394)NFYYAAVPSARNVPYGNIQSRLELRKKLSCKPFKWYLENVYPELRVPDHQ
    Consensus   (401)NFFY   P    V YG I SRL LR KL CKPF WYLENIYPD  IP H
                   451                                                 500
HSGALNAT1.pep   (429)FSLGEIRNVEINQCLDNMARKENEKVGIFNCHGMGGNQVFSYTANKEIRT
HSGALNAT2.pep   (444)DIAFGALQQGIN-CLDTLGHFADGVVGVYECHNAGGNQEWALTKEKSVKH
    Consensus   (451)      N  TN CLD LA      VGIF CH  GGNQ FA T  K IK
                   501                                                 550
HSGALNAT1.pep   (479)DDLCLDVSKLN--GPVIMLKCHHLKGNQLWEYDPVKLTLQHVNSNQCLDK
HSGALNAT2.pep   (493)MDLCLTVVDRAPGSLIKLQGCRENDSRQKWEQIEGNSKLRHVGSNLCLDS
    Consensus   (501) DLCL V        I L  C      Q WE       L HV SN CLD
                   551                 .      584
HSGALNAT1.pep   (527)ATEEDSQVPSIRDCNGSRSQQWLLRNVTLPEIF-
HSGALNAT2.pep   (543)RTAK-SGGLSVEVCGPALSQQWKFTLNLQQ----
    Consensus   (551)T   S   SI  C  A SQQW
```

FIG. 44

[0061] Figure 5 provides the results of an assay of GlycoPEGylation of EPO using the refolded SuperGlycoMix. Lanes are as follows: (1) MW markers, SeeBlue2 Invitrogen,(250, 148, 98, 64, 50, 36, 22, 16, 6 kD); (2) Positive control with EPO, + NSO expressed GalT1, BV GnT1, Aspergillus ST3GalIII and sugar nucleotides; (3) Negative control, Same as 2 without UDP-GlcNAc; (4) EPO, Purified and separately refolded MBP-GalT1(Δ129) C342T, Refolded MBP-GnT1(Δ103), and *Aspergillus niger* expressed ST3GalIII; (5) EPO, SuperGlycoMix (mixture of MBP-ST3GalIII, MBP-GalT1(Δ129) C342T, MBP-GnT1(Δ103)C123A and sugar nucleotides.

[0062] Figure 6 provides an alignment of a human GnT1 amino acid sequence (top line, ; SEQ ID NO:1 NP_002397) and a rabbit GnT1 amino acid sequence (bottom line, P27115). ; SEQ ID NO:2 The conserved unpaired cysteines are underlined and in bold text.

[0063] Figure 7 provides the amino acid sequence (SEQ ID NO:3) of a GnT1 Cys121Ser mutant and a nucleic acid sequence (SEQ ID NO:4) that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrrsldkllh.... (SEQ ID NO:5) where the bold residue is mutated from the wild-type cysteine.

[0064] Figure 8 provides the amino acid sequence (SEQ ID NO:6) of a GnT1 Cys121Asp mutant and a nucleic acid sequence (SEQ ID NO:7) that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrrdldkllh... (SEQ ID NO:8) where the bold residue is mutated from the wild-type cysteine.

[0065] Figure 9 provides the amino acid sequence (SEQ ID NO:9) of a GnT1 Cys121Thr mutant and a nucleic acid sequence (SEQ ID NO:10) that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrrtldkllh... (SEQ ID NO:11) where the bold residue is mutated from the wild-type cysteine.

[0066] Figure 10 provides the amino acid sequence (SEQ ID NO:12) of a GnT1 Cys121Ala mutant and a nucleic acid sequence (SEQ ID NO:13) that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrraldkllh... (SEQ ID NO:14) where the bold residue is mutated from the wild-type cysteine.

[0067] Figure 11 provides the amino acid sequence (SEQ ID NO:15) of a GnT1 Arg120Ala, Cys121His mutant and a nucleic acid sequence (SEQ ID NO:16) that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following double mutation: ...stvrahldkllh... (SEQ ID NO:17), where the bold residue is mutated from the wild-type cysteine.

[0068] Figure 12 provides the amino acid sequence of rat liver ST3GalIII (SEQ ID NO:18). The underlined and italicized sequence was deleted to make the Δ28 deletion.

[0069] Figures 13A and 13B provide full length nucleic acid (SEQ ID NO:20) and amino acid (SEQ ID NO:19) sequences of UDP-N-acetylgalactosaminyltransferase 2 (GalNAcT2). The accession number of the nucleic acid and protein is NM_004481.

[0070] Figures 14A and 14B provide nucleic acid (SEQ ID NO:22) and amino acid (SEQ ID NO:21) sequences of a Δ51GalNAcT2. The numbering is based on the full length amino acid and nucleic acid sequences shown in Figures 13A and B.

[0071] Figure 15 provides a demonstration of the protein concentration of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The pH values tested are expressed as solubilization pH-refolding pH. Protein concentrations were measured immediately after refolding (light gray bars), after dialysis (dark gray bars), and after concentration (white bars).

[0072] Figure 16 provides a demonstration of the enzymatic activity of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The pH values tested are expressed as solubilization pH-refolding pH. Activity was measured after dialysis (light gray bars) and after concentration (dark gray bars).

[0073] Figure 17 provides a demonstration of the specific activity of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The pH values tested are expressed as solubilization pH-refolding pH. Specific activity was measured after dialysis (white bars) and after concentration (dark gray bars).

[0074] Figures 18A and 18B provide results of remodeling of recombinant granulocyte colony stimulating factor (GCSF) using refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. Figure 18A shows the results using a control purified MBP-GalNAcT2(D51), or a negative control that lacked a substrate, or

bacterially expressed MBP-GalNAcT2(D51) that was solubilized at pH 6.5 and refolded at pH 6.5. Figure 18B shows the experimental results.

[0075] Figure 19 provides a profile of refolded MBP-GalNAcT2(D51) proteins after elution from a Q Sepharose XL (QXL) column (Amersham Biosciences, Piscataway, NJ). The top of the figure shows a chromatogram illustrating the elution of MBP-GalNAcT2(D51) from the QXL column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis. The bottom shows an image of two electrophoretic gels used to visualize the eluted fractions. The contents of each lane on the gel are described in the figure.

[0076] Figure 20 provides the GalNAcT2 activity of specific column fractions from the QXL column shown in Figure 19. The most active fractions were applied to a Hydroxyapatite Type I (80μm) (BioRad, Hercules, CA) column.

[0077] Figure 21 provides a profile of refolded MBP-GalNAcT2(D51) proteins after elution from the HA type I column. The top of the figure shows a chromatogram illustrating the elution of MBP-GalNAcT2(D51) from the HA type I column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis. The bottom shows an image of an electrophoretic gel used to visualize the eluted fractions. The contents of each lane on the gel are described in the figure.

[0078] Figure 22 provides the GalNAcT2 activity of HA type I eluted fractions.

[0079] Figure 23 provides a comparison of purification and activity of ST3Gal3 proteins fused to either an MBP tag or to an MBP-SBD tag.

[0080] Figure 24 provides the amino acid sequences of the MBP-ST3Gal1 fusion protein (SEQ ID NO:23) (A) and the MBP-SBD-ST3Gal1 fusion protein (SEQ ID NO:24) (B).

[0081] Figure 25 provides the sialyltransferase activity of the MBP-ST3Gal3 fusion protein) and the MBP-SBD-ST3Gal3 fusion protein. positive and negative controls are also shown.

[0082] Figure 26 provides the amino acid sequence of mouse and human ST6GalNAcI proteins fused to MBP. Part A shows the sequence of a mouse truncation fusion: MBP-mST6GalNAcI S127 (SEQ ID NO:25). Part B shows the sequence of a human truncation fusion: MBP-hST6GalNAcI K36 (SEQ ID NO:26)

[0083] Figure 27 provides SDS-PAGE gels of O-linked glycosyltransferase enzyme (A) concentrations after co-refolding and the (B) results of an enzyme assay after co-refolding. MBP-GalNAcT2 and MBP-ST3GalI were co-refolded together. Enzyme activity was tested after addition of Core I Gal T1 enzyme. The substrates were IFα-2b and 20K-Peg-CMP-NAN.

[0084] Figure 28 provides an SDS-PAGE gel showing expression of the native SiaA protein in *E. coli* before and after induction with IPTG.

[0085] Figure 29 provides an SDS-PAGE gel showing expression of an MBP-SiaA fusion protein in *E. coli* before and after induction with IPTG.

[0086] Figure 30 provides the amino acid sequence of the full length bovine GalT1 protein. (SEQ ID NO:27)

[0087] Figure 31 depicts GalT1 mutants schematically, as well as a control protein GalT1(40) (S96A+C342T).

[0088] Figure 32 provides the results of enzymatic assays of the refolded and purified MBP-GalT1 (D70) protein. The assay measured conversion of LNT2 (Lacto-N-Triose-2) into LNnT (Lacto-N-Neotetraose) using UDP-Gal (Uridine 5'-Diphosphogalactose) as a donor substrate.

[0089] Figure 33 provides an RNAse B remodeling assay of MBP-GalT1 (D70) and a control protein GalT1(40) (S96A+C342T), also referred to as Qasba's GalT1.

[0090] Figure 34 provides kinetics of glycosylation of RNAse B using the refolded and purified MBP-GalT1 (D70) protein or NSO GalT1, a soluble form of the bovine GalT1 protein that was expressed in a mammalian cell system.

[0091] Figure 35 provides a schematic of the MBP-GnT1 fusion proteins, and depicts the truncations, *e.g.*, Δ103 or Δ35, and the Cys121Ser mutation (top). The bottom of the figure provides the full length human GnT1 protein. (SEQ ID NO:1)

[0092] Figure 36 provides an SDS-PAGE gel showing in the right panel the refolded MBP-GnT1 fusion proteins: MBP-GnT1(D35) C121A, MBP-GnT1(D103) R120A + C121H, and MBP-GnT1(D103) C121A. The left panel shows GnT1 activities of two different batches (A1 and A2) of refolded MBP-GnT1(D35) C121A at different time points.

[0093] Figure 37 provides a full length sequence of porcine ST3Gal1. (SEQ ID NO:28)

[0094] Figure 38 provides full length amino acid sequences for A) human ST6GalNAcI (SEQ ID NO:29) and for B) chicken ST6GalNAcI (SEQ ID NO:30) and C) a sequence of the mouse ST6GalNAcI protein beginning at residue 32 of the native mouse protein (SEQ ID NO:31).

[0095] Figure 39 provides a schematic of a number of preferred human ST6GalNAcI truncation mutants.

[0096] Figure 40 shows a schematic of MBP fusion proteins including the human ST6GalNAcI truncation mutants.

[0097] Figure 41 provides the full length sequence of human Core 1 GalT1 protein (SEQ ID NO:32).

[0098] Figure 42 provides the sequences of two Drosophila Core 1 GalT1 proteins (SEQ ID NOS:33 and 34).

[0099] Figure 43 provides the sequences of exemplary bacterial MBP proteins that can be fused to glycosyltransferases to enhance refolding. A. *Yersinia* MBP (SEQ ID NO:35); B. *E. coli* MBP (SEQ ID NO:36); C. *Pyrococcus furiosus* MBP (SEQ ID NO:37); D. *Thermococcus litoralis* MBP (SEQ ID NO:38); E. *Thermatoga maritime* MBP (SEQ ID NO:39); and F. *Vibrio cholerae* MBP (SEQ ID NO:40).

[0100] Figure 44 provides an alignment of human GalNAcT1 (SEQ ID NO:41) and GalNAcT2 proteins (SEQ ID NO:19). Because the alignment programs account for sequence insertions or deletions, the numbering of cysteine residues is not the same as mentioned text and published sequences. In the case of hGalNAc-T2 cysteine 227 (published) corresponds to position 235 in the alignment and cysteine 229 (published) is 237 in the alignment. The hGalNAc-T1 cysteines are 212 (published), which corresponds to cysteine 235 (alignment) and 214 (published), which corresponds to cysteine 237 (alignment). The relevant cysteine residues are indicated by larger font size. Consensus peptides = SEQ ID NOS:42-65.

[0101] Figure 45 shows the position of paired and unpaired cysteine residues in the human ST6GalNAcI protein. Single and double cysteine substitution are also shown, *e.g.*, C280S, C362S, C362T, (C280S + C362S), and (C280S + C362T).

DEFINITIONS

[0102] The recombinant glycosyltransferase proteins of the invention are useful for transferring a saccharide from a donor substrate to an acceptor substrate. The addition generally takes place at the non-reducing end of an oligosaccharide or carbohydrate moiety on a biomolecule. Biomolecules as defined here include but are not limited to biologically

[0133] A "fusion protein" refers to a protein comprising amino acid sequences that are in addition to, in place of, less than, and/or different from the amino acid sequences encoding the original or native full-length protein or subsequences thereof.

[0134] Components of fusion proteins include "accessory enzymes" and/or "purification tags." An "accessory enzyme" as referred to herein, is an enzyme that is involved in catalyzing a reaction that, for example, forms a substrate for a glycosyltransferase. An accessory enzyme can, for example, catalyze the formation of a nucleotide sugar that is used as a donor moiety by a glycosyltransferase. An accessory enzyme can also be one that is used in the generation of a nucleotide triphosphate required for formation of a nucleotide sugar, or in the generation of the sugar which is incorporated into the nucleotide sugar. The recombinant fusion protein of the invention can be constructed and expressed as a fusion protein with a molecular "purification tag" at one end, which facilitates purification of the protein. Such tags can also be used for immobilization of a protein of interest during the glycosylation reaction. Suitable tags include "epitope tags," which are a protein sequence that is specifically recognized by an antibody. Epitope tags are generally incorporated into fusion proteins to enable the use of a readily available antibody to unambiguously detect or isolate the fusion protein. A "FLAG tag" is a commonly used epitope tag, specifically recognized by a monoclonal anti-FLAG antibody, consisting of the sequence AspTyrLysAspAspAspAspLys (SEQ ID NO:66) or a substantially identical variant thereof. Other suitable tags are known to those of skill in the art, and include, for example, an affinity tag such as a hexahistidine peptide (SEQ ID NO:67), which will bind to metal ions such as nickel or cobalt ions. Proteins comprising purification tags can be purified using a binding partner that binds the purification tag, *e.g.*, antibodies to the purification tag, nickel or cobalt ions or resins, and amylose, maltose, or a cyclodextrin. Purification tags also include starch binding domains, *E. coli* thioredoxin domains (vectors and antibodies commercially available from *e.g.*, Santa Cruz Biotechnology, Inc. and Alpha Diagnostic International, Inc.), and the carboxy-terminal half of the SUMO protein (vectors and antibodies commercially available from *e.g.*, Life Sensors Inc.). Maltose binding domains are preferably used for their ability to enhance refolding of insoluble eukaryotic glycosyltransferases, but can also be used to assist in purification of a fusion protein. Purification of maltose binding domain proteins is known to those of skill in the art. Starch binding domains are described in WO 99/15636, herein incorporated by reference. Affinity purification of a fusion protein comprising a starch binding domain using

about amino acid residues 32-90. Thus, a truncated human Core1 GalT1 protein can have deletions at the amino terminus of about *e.g.*, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90 residues.

[0210] Deletion mutations can also be made in an ST3Gal1 protein. For example, the human ST3Gal1 protein includes a stem region from about amino acid residues 18-58. Thus, a truncated human ST3Gal1 protein can have deletions at the amino terminus of about *e.g.*, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, or 58 residues. As another example, the porcine ST3Gal1 protein includes a stem region from about amino acid residues 28-61. Thus, a truncated porcine ST3Gal1 protein can have deletions at the amino terminus of about *e.g.*, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, or 61 residues.

[0211] Deletion mutations can also be made in a GalNAcT2 protein. For example, the rat GalNAcT2 protein includes a stem region from about amino acid residues 40-95. Thus, a truncated rat GalNAcT2 protein can have deletions at the amino terminus of about *e.g.*, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, or 95 residues.

[0212] Deletion mutations can also be made in an ST6GalNAcI protein. For example, the mouse ST6GalNAcI protein includes a stem region from about amino acid residues 30-207. Thus, a truncated mouse ST6GalNAcI protein can have deletions at the amino terminus of about *e.g.*, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197 198, 199, 200, 201, 202, 203, 204, 205, 206, or 207 residues. As another example, the human ST6GalNAcI protein includes a stem region from about amino

acid residues 35-278. Thus, a truncated human ST6GalNAcI protein can have deletions at the amino terminus of about *e.g.*, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277,or 278 residues. As still another example, chicken ST6GalNAcI protein includes a stem region from about amino acid residues 37-253. Thus, a truncated chicken ST6GalNAcI protein can have deletions at the amino terminus of about *e.g.*, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, or 253 residues.

D. One pot refolding of glycosyltransferases

[0213] These embodiments of the invention are based on the surprising observation that multiple eukaryotic glycosyltransferases expressed in bacterial inclusion bodies can be refolded in a single vessel, *i.e.*, a one pot method. Using this method at least two glycosyltransferases can be refolded together resulting in savings of time and materials.

residues of the yeast ubiquitin gene containing a peptidase cleavage site. Cleavage at the junction of the two moieties results in production of a protein having an intact authentic N-terminal reside.

[0259] The expression vectors of the invention can be transferred into the chosen host cell by well-known methods such as calcium chloride transformation for *E. coli* and calcium phosphate treatment or electroporation for mammalian cells. Cells transformed by the plasmids can be selected by resistance to antibiotics conferred by genes contained on the plasmids, such as the *amp, gpt, neo* and *hyg* genes.

VI. Proteins and protein purification

[0260] The recombinant eukaryotic glycosyltransferase proteins can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (*see*, generally, R. Scopes, *Protein Purification*, Springer-Verlag, N.Y. (1982), Deutscher, *Methods in Enzymology Vol. 182: Guide to Protein Purification.*, Academic Press, Inc. N.Y. (1990)). In preferred embodiments, purification of the recombinant eukaryotic glycosyltransferase proteins occurs after refolding of the protein. Substantially pure compositions of at least about 70 to 90%, homogeneity are preferred; more preferably at least 91%, 92%, 93%, 94%, 95%, 96%, or 97%; and 98 to 99% or more homogeneity are most preferred. The purified proteins may also be used, *e.g.*, as immunogens for antibody production.

[0261] To facilitate purification of the recombinant eukaryotic glycosyltransferase proteins of the invention, the nucleic acids that encode the recombinant eukaryotic glycosyltransferase proteins can also include a coding sequence for an epitope or "tag" for which an affinity binding reagent is available, *i.e.* a purification tag. Examples of suitable epitopes include the myc and V-5 reporter genes; expression vectors useful for recombinant production of fusion proteins having these epitopes are commercially available (*e.g.*, Invitrogen (Carlsbad CA) vectors pcDNA3.1/Myc-His and pcDNA3.1/V5-His are suitable for expression in mammalian cells). Additional expression vectors suitable for attaching a tag to the fusion proteins of the invention, and corresponding detection systems are known to those of skill in the art, and several are commercially available (e.g., "FLAG" (Kodak, Rochester NY). Another example of a suitable tag is a polyhistidine sequence, which is capable of binding to metal chelate affinity ligands. Typically, six adjacent histidines (SEQ ID NO:67) are used, although one can use more or less than six. Suitable metal chelate affinity ligands that can serve as the binding

enzyme, GST-ST3-GalIII, was active and transferred sialic acid to an LNnT sugar substrate and to asialylated glycoproteins, for example, transferrin and Factor IX.

Cloning ST3GalIII into pGEX-XT-KT vector

[0305] Rat liver ST3-GalIII gene was cloned into *Bam*H1 and *Eco*R1 sites of the pGEX-KT-Ext vector after PCR Amplification using the following primers:

Sense Sial 5'Tm 5'-TTTGGATCCAAGCTACACTTACTCCAATGG (SEQ ID NO:68)
Antisense: Sial 3' Whole 5'-TTTGAATTCTCAGATACCACTGCTTAAGTC (SEQ ID NO:69)

Expression of GST-ST3GalIII in *E. coli* BL21cells

[0306] pGEX-ST3GalIII, an expression vector comprising the ST3GalIII GST fusion, was transformed into chemically competent *E. coli* BL21 cells. Single colonies were picked, inoculated into five ml LB media with 100 µg/ml carbenicillin, and grown overnight at 37°C with shaking. The next day, one ml of overnight culture was transferred into one liter of LB media with 100 µg/ml carbenicillin. Bacteria were grown until to an OD_{620} of 0.7, then 150 µM IPTG (final) was added to the medium. Bacteria were grown at 37°C for one to two hours more, then shifted to room temperature and grown overnight with shaking. Cells were harvested by centrifugation; bacterial pellets were resuspended in PBS buffer and lysed using a French Press. Soluble and insoluble fractions were separated by centrifugation for thirty minutes at 10,000 RPM in a Sorvall, SS 34 rotor at 4°C.

Purification of the inclusion bodies

[0307] Fifty ml of Novagen's Wash buffer (20 mM Tris.HCl, pH 7.5, 10 mM EDTA, 1 % Triton X-100) was added to the insoluble fraction, *i.e.*, the inclusion bodies (IB's). The insoluble fraction was vortexed to resuspend the pellet. The suspended IB's were centrifuged and washed at least twice by resuspending in Wash Buffer as above. Clean precipitates (IB's) were recovered and were stored at -20 °C until use.

Refolding inclusion bodies

[0308] The IB's were weighed (144 mg) and dissolved in Genotech IBS buffer (1.44 ml). The resuspended IB's were incubated at 4 °C for one hour in an Eppendorf centrifuge tube. Insoluble material was removed by centrifugation at maximum speed in an Eppendorf centrifuge. Solubilized IB's were diluted to 4 ml final volume. Refolding of GST-ST3GalIII was tested in refolding buffer solutions containing cyclodextrin, polyethylene glycol (PEG), ND SB-201, or a GSH/GSSG redox system. One ml of solubilized IB's were diluted rapidly by pipetting into the refolding solution, vigorously mixed for 30-40 seconds, and then gently

Table 3. GST-ST3GalIII activities after two separate folding experiments using GSH/GSSG system.

GSH/GSSG	Conc	Activity
Refolding Trial 1	12x	182 U/L*
Refolding Trial 2	40x	531 U/L*

*Activities reported here are Units per L refolded enzyme

Sialylation of glycoproteins using refolded GST-ST3 Gal III

[0311] Twenty μL of asialylated Transferrin (2μg/μL) or asialylated Factor IX (2μg/μL), was added to fifty μL of a buffer containing 50mM Tris, pH 8.0; and 150 mM NaCl, with 10 μL of 100 mM $MnCl_2$; 10 μL of 200mM CMP-NAN; and 0.05% sodium azide. The reaction mixture was incubated with 30 μL refolded GST-ST3GalIII at 30°C overnight or longer with shaking at 250 rpm. After the reactions were stopped, the sialylated proteins were separated on pH 7-3 IEF (Isoelectric focusing gel, Invitrogen) and stained with Comassie Blue according to manufacturer's guideline. Both Transferrin and Factor IX were sialylated by GST-ST3GalIII. (Data not shown.)

Refolding a rat liver ST3GalIII fused to an MBP tag.

[0312] Rat liver ST3GalIII was cloned into pMAL-c2x vector and expressed as a maltose binding protein (MBP) fusion, MBP-ST3GalIII, in inclusion bodies of *E.coli* TB1 cells. The refolded MBP-ST3GalIII was active and transferred sialic acid to LNnT, a sugar substrate, and to asialylated glycoproteins, for example asialo-transferrin.

Cloning ST3GalIII into pMAL-c2x vector

[0313] The rat liver ST3-GalIII nucleic acid was cloned into *Bam*H1 and *Xba*I sites of the pMAL-c2x vector after PCR Amplification using the following primers:

Sense ST3BAMH1 5'-TAATGGATTCAAGCTACACTTACTCCAATGG (SEQ ID NO:70)
Antisense: ST3XBA1 5'-GCGCTCTAGATCAGATACCACTGCTTAAGT (SEQ ID NO:71)

[0314] Nucleotides encoding amino acids 28-374, *e.g.*, the stem region and catalytic domain of ST3GalIII, were fused to the MBP amino acid tag.

[0315] Three other truncations of ST3GalIII were constructed and fused to MBP. The three ST3Gal III (Δ73, Δ85, Δ86) inserts were isolated by PCR using the following 5' primers (ST3 BamH1 Δ73) TGTATCGGATCCCTGGCCACCAAGTACGCTAACTT (SEQ ID NO:72); (ST3 BamH1 Δ85) TGTATCGGATCCTGCAAACCCGGCTACGCTTCAGCCAT (SEQ ID NO:73); and (ST3

BamH1 Δ86) TGTATCGGATCCAAACCCGGCTACGCTTCAGCCAT (SEQ ID NO:74) respectively, in pairs with the common 3' primer (ST3-XhoI) GGTCTCCTCGAGTCAGATACCACTGCTTAA (SEQ ID NO:75). Each PCR product was digested with BamHI and XhoI, subcloned into BamHI-XhoI digested pCWin2-MBP Kanr vector, transformed into TB1 cells, and screened for the correct construct.

[0316] PCR reactions were carried out under the following conditions. One cycle at 95°C for 1 minute. One µl vent polymerase was added. Ten of the following cycles were performed: 94°C for 1 minute; 65°C for 1 minute; and 72°C for 1 minute. After a final ten minutes at 72°C, the reaction was cooled to 4°C.

[0317] All of the ST3GalIII truncations had activity after refolding. The experiments described below were performed using the MBP Δ73ST3GalIII truncation.

Expression of MBP-ST3GalIII in *E. coli* TB1 cells

[0318] The pMAL-ST3GalIII plasmid was transformed into chemically competent *E. coli* TB1 cells. Three isolated colonies containing TB1/pMAL-ST3GalIII construct were picked from the LB agar plates. The colonies were grown in five ml of LB media supplemented with 60 µg/ml carbenicillin at 37°C with shaking until the liquid cultures reached an OD_{620} of 0.7. Two one ml aliquots were withdrawn from each culture and used to inoculate fresh media with or without 500 µM IPTG (final). The cultures were grown at 37°C for two hours. Bacterial cells were harvested by centrifugation. Total cell lysates were prepared heating the cell pellets in the presence of SDS and DTT. IPTG induced expression of MBP-ST3GalIII. (Data not shown.)

Expression of MBP-ST3GalIII and Purification of the inclusion bodies:

[0319] A one ml aliquot of TB1/pMAL-ST3GalIII overnight culture was inoculated into 0.5 liter of LB media with 50 µg/ml carbenicillin and grown to an OD_{620} of 0.7. Expression of MBP-ST3GalIII was induced by addition of 0.5 mM IPTG, followed by overnight incubation at room temperature. The next day bacterial cells were harvested by centrifugation. Cell pellets were resuspended in a buffer containing 75 mM TrisHCl, pH 7.4; 100 mM NaCl; and 1 % glycerol. Bacterial cells were lyzed using a French Press. Soluble and insoluble fractions were separated by centrifugation for thirty minutes, 4°C, 10,000 rpm, Sorvall, SS 34 rotor). Soluble and insoluble fractions were separated by centrifugation for thirty minutes at 10,000RPM in a Sorvall, SS 34 rotor at 4°C.

buffer was supplemented with 0.3 mM Lauryl maltoside (LM); 0.1 mM oxidized glutathione (GSSG); 1 mM reduced glutathione (GSH) immediately before the addition of solubilized IB's. Two ml of solubilized IB's were added into 43 ml of refolding buffer in 50 ml sterile culture tube. The tube was placed on a rocker-shaker and gently shaken for 24 hours at 4°C. The refolded protein was dialyzed in dialysis tubing (MWCO: 7 kD) against Dialysis Buffer (100 mM Tris HCl, pH 7.5; 100 mM NaCl; and 5 % glycerol) twice (in 10-20 volume excess buffer).

[0330] The large scale dialyzed, refolded MBP-Gal III was analyzed for ST3GalIII activity, and exhibited about 53.6 U/g IB.

Example 2: Site Directed Mutagenesis of Human GnTI to Enhance Refolding.

[0331] A truncated human N-acetylglucosaminyltransferase I (103 amino terminal amino acids deleted) was expressed in *E.coli* as a maltose binding fusion protein (GnTI/MBP). The fusion protein was insoluble and was expressed in inclusion bodies. After solubilization and refolding, the GnTI/MBP fusion protein had low activity. The crystal structure of a truncated form of rabbit GnTI (105 amino terminal amino acids deleted) shows an unpaired cysteine residue (CYS123) near the active site. (See, *e.g.*, Unligil *et al.*, *EMBO J.* 19:5269-5280 (2000)). The corresponding unpaired cysteine in the human GnTI was identified as CYS121 and was replaced with a series of amino acids that are similar in size and chemical characteristics. The amino acids used include serine (Ser), threonine (Thr), alanine (Ala) and aspartic acid (Asp). In addition, a double mutant, ARG120ALA, CYS121HIS, was also made. The mutant GnTI/MBP fusion proteins were expressed in *E. coli*, refolded and assayed for GnTI activity towards glycoproteins.

[0332] Mutagenesis was done using a Quick Change Site-Directed Mutagenesis Kit from Stratagene. Additional restriction sites were introduced with some of the GnT1 mutations. For example an *Apa*I site (underlined, GGGCCCAC) was introduced into the GnT1 ARG120ALA, CYS121HIS mutant, *i.e.*, CGC CTG → **GCC CAC** (changes in bold). The following mutagenic oligonucleotides were used to make the double mutant: GnT1 R120A, C121H+, 5'CCGCAGCACTGTTCGGGCCCACCTGGACAAGCTGCTG 3'; (SEQ ID NO:76) and GnT1 R120A, C121H- 5'CAGCAGCTTGTCCAGGTGGGCCCGAACAGTGCTGCGG 3'(SEQ ID NO:77) (changes shown in bold). An *Asc*I site (underlined, GGCGCGCC) was introduced into the GnT1 CYS121ALA mutant, *i.e.*, CTG → **GCC** (changes in bold). The following mutagenic oligonucleotides were used to make the GnT1 CYS121ALA mutant: GnT1C123A+

5'AGCACTGTTCGGCGCGCCCTGGACAAGCTGCTG 3' (SEQ ID NO:78) and GnT1C123A-
5'CAGCAGCTTGTCCAGGGCGCGCCGAACAGTGCT 3' (SEQ ID NO:79).

[0333] The activity of the mutant proteins expressed in *E. coli* was compared to the activity of wild type GnT1 expressed in baculovirus. A CYS121SER GNTI mutant was active in a TLC based assay. In contrast, a CYS121THR mutant had no detectable activity and a CYS121ASP mutant had low activity. A CYS121ALA mutant was very active, and a double mutant, ARG120ALA, CYS121HIS, based on the amino acid sequence of the *C. elegans* GnT1 protein (Gly14), also exhibited activity, including transfer of GlcNAc to glycoproteins. Amino acid and encoding nucleic acid sequences of the GnT1 mutants are provided in Figures 7-11.

[0334] A second GnT1 truncation was made and fused to MBP: MBP-GnT1(D35). Figure 35 provides a schematic of the MBP-GnT1 fusion proteins, and depicts the truncations, *e.g.*, Δ103 or Δ35, and the Cys121Ser mutation (top). The bottom of the figure provides the full length human GnT1 protein. Mutations of Cys121 were also made in the MBP-GnT1(D35) protein.

[0335] Both fusion proteins were expressed in E. coli and both had activity for remodeling of the RNAse B glycoprotein. Figure 36 provides an SDS-PAGE gel showing in the right panel the refolded MBP-GnT1 fusion proteins: MBP-GnT1(D35) C121A, MBP-GnT1(D103) R120A + C121H, and MBP-GnT1(D103) C121A. The left panel shows the activities for remodeling the RNAse B glycoprotein of two different batches (A1 and A2) of refolded MBP-GnT1(D35) C121A at different time points. The MBP-GnT1 (D103) C121A also remodeled the RNAse B glycoprotein. Data not shown.

Example 3: MPB fusions to GalT1.

[0336] The following fusions between truncated bovine GalT1 and MBP were constructed: MBP-GalT1 (D129) wt, (D70) wt or (D129 C342T). (For the full length bovine sequence, see, *e.g.*, D'Agostaro *et al.*, *Eur. J. Biochem.* 183:211-217 (1989) and accession number CAA32695.) Each construct had activity after refolding. The amino acid sequence of the full length bovine GalT1 protein is provided in Figure 30. The mutants are depicted shematically in Figure 31 with a control protein GalT1(40) (S96A+C342T). *See, e.g.*, Ramakrishnan *et al.*, *J. Biol. Chem.* 276:37666-37671 (2001).

[0337] MBP-GalT1 (D70) was expressed in *E. coli* strain JM109. After overnight induction with IPTG, inclusion bodies were isolated from the insoluble pellet after cells were

Example 5: Refolding eukaryotic GalNAcT2.

[0365] A truncated human GalNAcT2 enzyme was expressed in *E. coli* and used to determine optimal conditions for solubilization and refolding using the methods described above. The full length human GalNAcT2 nucleic acid and amino acid sequences are provided in Figures 13A and B. The sequences of the mutant protein, GalNAcT2(D51), are shown in Figures 14A and B. The mutant was expressed in *E. coli* as an MBP fusion protein, MBP-GalNAcT2(D51). Other GalNAcT2 mutants were made, expressed in E. coli and were able to be refolded: MBP-GalNAcT2(D40), MBP-GalNAcT2(D73), and MBP-GalNAcT2(D94). Data not shown. Details of the construction of the additional deletion mutants is found in USSN 60/576,530, filed June 3, 2004 and USSN 60/598,584, August 3, 2004, both of which are herein incorporated by reference for all purposes.

[0366] Cultures of bacteria expressing MBP-GalNAcT2(D51) were grown and harvested as described above. Inclusion bodies were purified from bacteria as described above. Solubilization of the inclusion bodies was performed at pH 6.5 or at pH 8.0. After solubilization, MBP-GalNAcT2(D51) protein was refolded at either pH 6.5 or pH 8.0 using buffers A and B, *i.e.*, **Buffer A**: 55 mM MES pH 6.5, 550 mM Arginine, 0.055 % PEG3350, 264 mM NaCl, 11 mM KCl, supplemented with 1 mM GSH, 0.1 mM GSSG; and **Buffer B**: 55 mM TrisHCl pH 8, 550 mM Arginine, 0.055 % PEG3350, 264 mM NaCl, 11 mM KCl, supplemented with 1 mM GSH, 0.1 mM GSSG. After refolding, MBP-GalNAcT2(D51) protein was dialyzed and then concentrated. Figure 15 provides a demonstration of the protein concentration of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0.

[0367] A radiolabeled [^{3}H]-UDP-GalNAc assay was performed to determine the activity of the *E.coli*-expressed refolded MBP-GalNAcT2(D51) by monitoring the addition of radiolabeled GalNAc to a peptide acceptor. The acceptor was a MuC-2 – like peptide having the sequence MVTPTPTPTC (SEQ ID NO:80). The peptide was dissolved in 1M Tris-HCl pH=8.0. See, *e.g.*, USSN 60/576,530 filed June 3, 2004; and US provisional patent application Attorney Docket Number 040853-01-5149-P1, filed August 3, 2004; both of which are herein incorporated by reference for all purposes. Figure 16 provides a demonstration of the enzymatic activity of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. Figure 17 provides a demonstration of the specific activity of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The highest activity levels were observed with MBP-GalNAcT2(D51) that

Human ST6GalNAcI

MRSCLWRCRHLSQGVQWSLLLAVLVFFLFALPSFIKEPQTKPSRHQRTENIKERSLQS
LAKPKSQAPTRARRTTIYAEPVPENNALNTQTQPKAHTTGDRGKEANQAPPEEQDK
VPHTAQRAAWKSPEKEKTMVNTLSPRGQDAGMASGRTEAQSWKSQDTKTTQGNG
GQTRKLTASRTVSEKHQGKAATTAKTLIPKSQHRMLAPTGAVSTRTRQKGVTTAVIP
PKEKKPQATPPPAPFQSPTTQRNQRLKAANFKSEPRWDFEEKYSFEIGGLQTTCPDSV
KIKASKSLWLQKLFLPNLTLFLDSRHFNQSEWDRLEHFAPPFGFMELNYSLVQKVVT
RFPPVPQQQLLLASLPAGSLRCITCAVVGNGGILNNSHMGQEIDSHDYVFRLSGALIK
GYEQDVGTRTSFYGFTAFSLTQSLLILGNRGFKNVPLGKDVRYLHFLEGTRDYEWLE
ALLMNQTVMSKNLFWFRHRPQEAFREALHMDRYLLLHPDFLRYMKNRFLRSKTLD
GAHWRIYRPTTGALLLLTALQLCDQVSAYGFITEGHERFSDHYYDTSWKRLIFYINH
DFKLEREVWKRLHDEGIIRLYQRPGPGTAKAKN

FIG. 38A

Chicken ST6GalNAcI

MGFLIRRLPKDSRIFRWLLILTVFSFIITSFSALFGMEKSIFRQLKIYQSIAHMLQVDTQ
DQQGSNYSANGRISKVGLERDIAWLELNTAVSTPSGEGKEEQKKTVKPVAKVEEAK
EKVTVKPFPEVMGITNTTASTASVVERTKEKTTARPVPGVGEADGKRTTIALPSMKE
DKEKATVKPSFGMKVAHANSTSKDKPKAEEPPASVKAIRPVTQAATVTEKKKLRAA
DFKTEPQWDFDDEYILDSSSPVSTCSESVRAKAAKSDWLRDLFLPNITLFIDKSYFNV
SEWDRLEHFAPPYGFMELNYSLVEEVMSRLPPNPHQQLLLANSSSNVSTCISCAVVG
NGGILNNSGMGQEIDSHDYVFRVSGAVIKGYEKDVGTKTSFYGFTAYSLVSSLQNLG
HKGFKKIPQGKHIRYIHFLEAVRDYEWLKALLLDKDIRKGFLNYYGRRPRERFDEDF
TMNKYLVAHPDFLRYLKNRFLKSKNLQKPYWRLYRPTTGALLLLTALHLCDRVSAY
GYITEGHQKYSDHYYDKEWKRLVFYVNHDFNLEKQVWKRLHDENIMKLYQRS

FIG. 38B

Mouse ST6GalNAcI protein beginning at residue 32 of the native mouse protein

DPRAKDSRCQFIWKNDASAQENQQKAEPQVPIMTLSPRVHNKESTSVSSKDLKKQER
EAVQGEQAEGKEKRKLETIRPAPENPQSKAEPAAKTPVSEHLDKLPRTPGALSTRKTP
MATGAVPAKKKVVQATKSPASSPHPTTRRRQRLKASEFKSEPRWDFEEEYSLDMSSL
QTNCSASVKIKASKSPWLQNIFLPNITLFLDSGRFTQSEWNRLEHFAPPFGFMELNQSL
VQKVVTRFPPVRQQQLLLASLPTGYSKCITCAVVGNGGILNDSRVGREIDSHDYVFR
LSGAVIKGYEQDVGTRTSFYGFTAFSLTQSILILGRRGFQHVPLGKDVRYLHFLEGTR
NYEWLEAMFLNQTLAKTHLSWFRHRPQEAFRNALDLDRYLLLHPDFLRYMKNRFL
RSKTLDTAHWRIYRPTTGALLLLTALHLCDKVSAYGFITEGHQRFSDHYYDTSWKRL
IFYINHDFRLERMVWKRLHDEGIIWLYQRPQSDKAKN

FIG. 38C

```
                    1                                                  50
HSGALNAT1.pep     (1)MRKFAYCKVVLATSLIWVLLDMFLLLYFS----------------ECNKC
HSGALNAT2.pep     (1)MRRRS--RMLLCFAFLWVLGIAYYMYSGGGSALAGGAGGGAGRKEDWNEI
    Consensus     (1)MRK A  KMLL  A IWVL   F L                     D N
                    51                                                 100
HSGALNAT1.pep    (35)DEKKERGLPAGDVLEPVQKPHEGP-G-----EMGKPVVIPKEDQEKMKEM
HSGALNAT2.pep    (49)DPIKKKDLHHSNGEEKAQSMETLPPGKVRWPDFNQEAYVGGIMVRSGQDP
    Consensus    (51)D  K K L      E  Q     P G     D      I          D
                   101                                                 150
HSGALNAT1.pep    (79)FKINQFNLMASEMIALNRSLPDVRLEGCKTKVYPDNLPTTSVVIVFHNEA
HSGALNAT2.pep    (99)YARNKFNQVESDKLRMDRAIPDTRHDQCQRKQWRVDLPATSVVITFHNEA
    Consensus   (101)F  N FN M SD I L RAIPD R D C  K W   LP TSVVI FHNEA
                   151                                                 200
HSGALNAT1.pep   (129)WSTLLRTVHSVINRSPRHMIEEIVLVDDASERDFLKRPLESYVKKLKVPV
HSGALNAT2.pep   (149)RSALLRTVVSVLKKSPPHLIKEIILVDDYSN-----DPEDGALLGKIEKV
    Consensus   (151) S LLRTV SVI KSP HLI EIILVDD S       P D  L      V
                   201                                                 250
HSGALNAT1.pep   (179)HVIRMEQRSGLIRARLKGAAVSKGQVITFLDAHCECQTVGWLEPLLARIKH
HSGALNAT2.pep   (194)RVLRNDRREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAE
    Consensus   (201) VIR D R GLIRARLKGA  A A VITFLDAHCEC   WLEPLL RI
                   251                                                 300
HSGALNAT1.pep   (229)DRRTVVCPIIDVISDDTFEYMAGSDMTYGGFNWKLNFRWYPVPQREMDRR
HSGALNAT2.pep   (244)DRTRVVSPIIDVINMDNFQYVGASADLKGGFDWNLVFKWDYMTPEQRRSR
    Consensus   (251)DR  VV PIIDVI  D F YMAAS    GGF W L FKW  M       R
                   301                                                 350
HSGALNAT1.pep   (279)KGDRTLPVRTPIMAGGLFSIDRDYFQEIGTYDAGMDIWGGENLEISFRIW
HSGALNAT2.pep   (294)QGNPVAPIKTPMIAGGLFVMDKFYFEELGKYDMMMDVWGGENLEISFRVW
    Consensus   (301) G    PIKTP IAGGLF IDK YF EIG YD  MDIWGGENLEISFRIW
                   351                                                 400
HSGALNAT1.pep   (329)QCGGTLEIVTCSHVGHVFRKATPYTFPGGTGQIINKNNRRLAEVWMDEFK
HSGALNAT2.pep   (344)QCGGSLEIIPCSRVGHVFRKQHPYTFPGGSGTVFARNTRRAAEVWMDEYK
    Consensus   (351)QCGGSLEII CS VGHVFRK  PYTFPGGSG I  KN RR AEVWMDEFK
                   401                                                 450
HSGALNAT1.pep   (379)NFFYIISPGVTKVDYGDISSRVGLRHKLQCKPFSWYLENIYPDSQIPRHY
HSGALNAT2.pep   (394)NFYYAAVPSARNVPYGNIQSRLELRKKLSCKPFKWYLENVYPELRVPDHQ
    Consensus   (401)NFFY    P    V YG I SRL LR KL CKPF WYLENIYPD  IP H
                   451                                                 500
HSGALNAT1.pep   (429)FSLGEIRNVETNQCLDNMARKENEKVGIFNCHGMGGNQVFSYTANKEIRT
HSGALNAT2.pep   (444)DIAFGALQQGIN-CLDILGHFADGVVGVYECHNAGGNQEWALTKEKSVKH
    Consensus   (451)       N  TN CLD LA      VGIF CH  GGNQ FA T  K IK
                   501                                                 550
HSGALNAT1.pep   (479)DDLCLDVSKLN--GPVIMLKCHHLKGNQLWEYDPVKLTLQHVNSNQCLDK
HSGALNAT2.pep   (493)MDLCLTVVDRAPGSLIKLQGCRENDSRQKWEQIEGNSKLRHVGSNLCLDS
    Consensus   (501) DLCL V        I L  C      Q WE       L HV SN CLD
                   551                                584
HSGALNAT1.pep   (527)ATEEDSQVPSIRDCNGSRSQQWLLRNVTLPEIF-
HSGALNAT2.pep   (543)RTAK-SGGLSVEVCGPALSQQWKFTLNLQQ----
    Consensus   (551) T   S   SI  C  A SQQW
```

Bold

Larger Font Size

FIG. 44